



Predicting Personalities



Determining Myer-Briggs Personality
Types From Online Social Postings



What your online
posts say about
you.

Using Natural Language
Processing and Recurrent
Neural Network Modeling, how
accurately can we identify a
person's personality type?

The Myers-Briggs Type Indicator

An introspective self- reporting
questionnaire.

The MBTI is based on conceptual
theory from psychiatrist Carl Jung.

People experience the world using
four principle psychological
functions:

- Sensation
 - Intuition
 - Feeling
 - Thinking
-

Each of these functions correspond with how a person approaches *energy, information, organization, and decisions*.

16 unique personality types result from variations in these approaches.

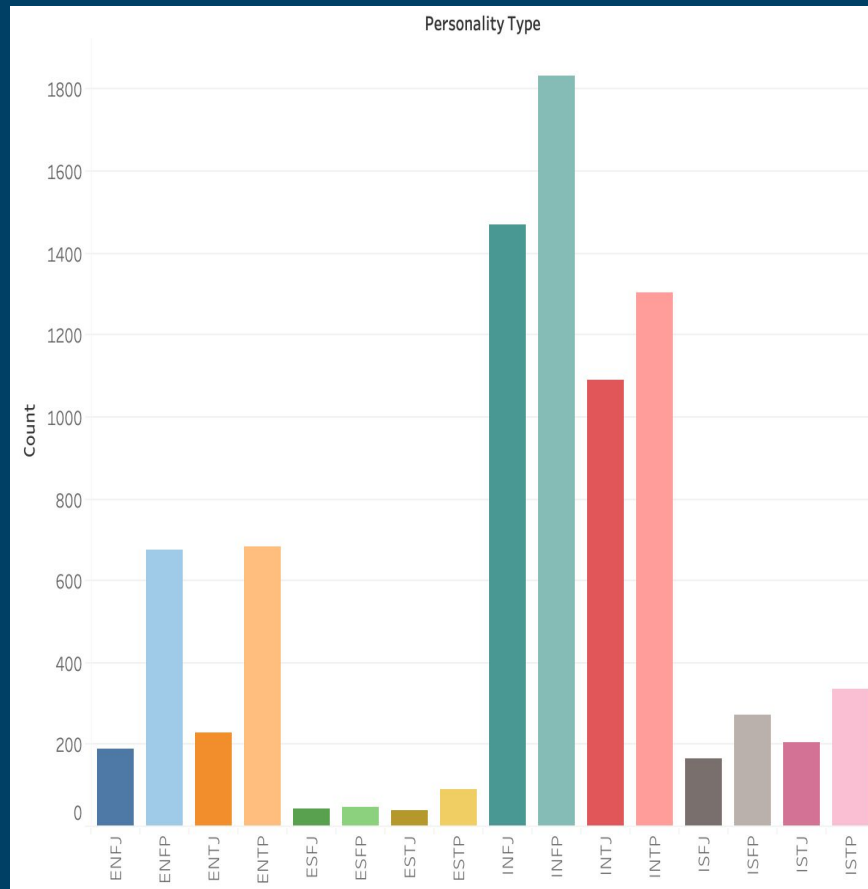


Looking at the Data

The Myers-Briggs Personality Dataset : Kaggle

Data came through the PersonalityCafe forum.

Consisted of 8,600 users, their MBTI personality type and their last 50 postings.



The Personality Traits

Are you outwardly or inwardly focused?

E

Extraversion

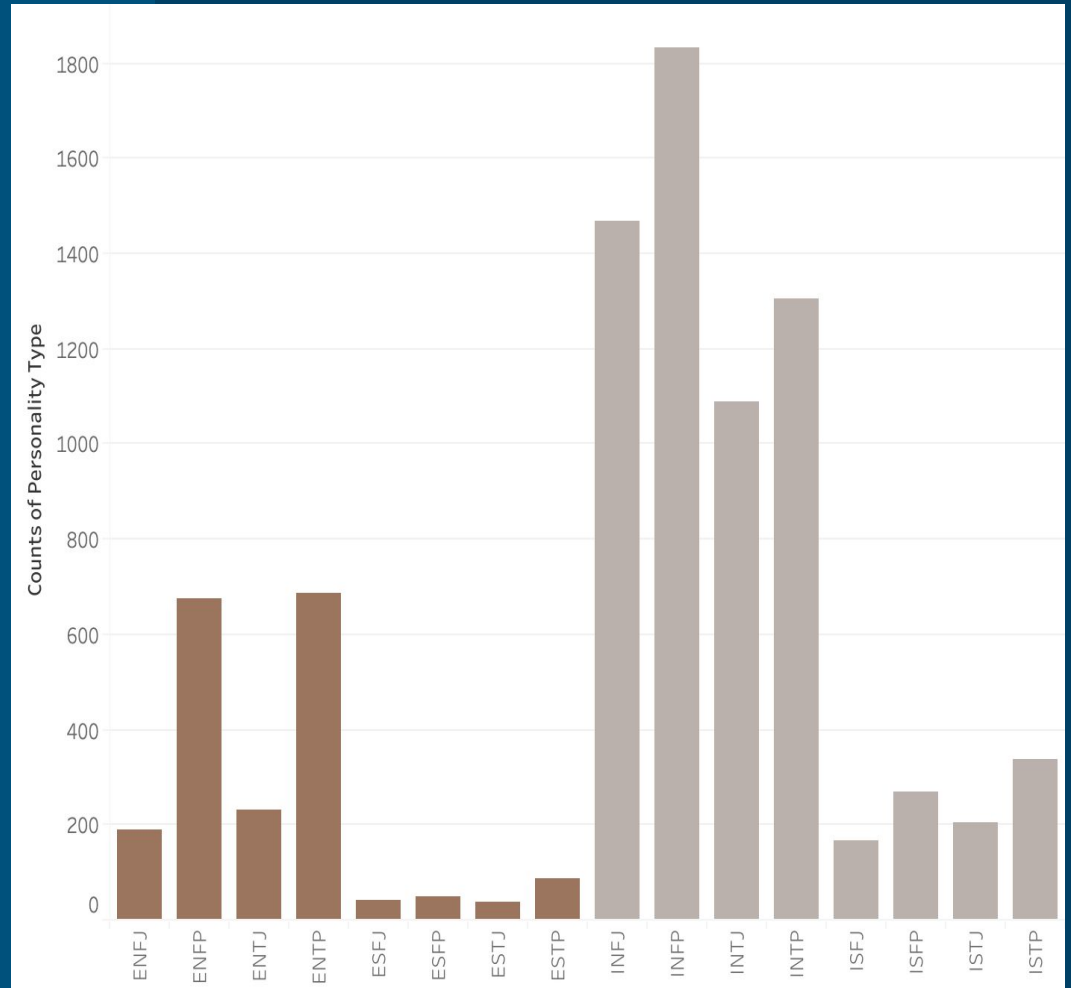
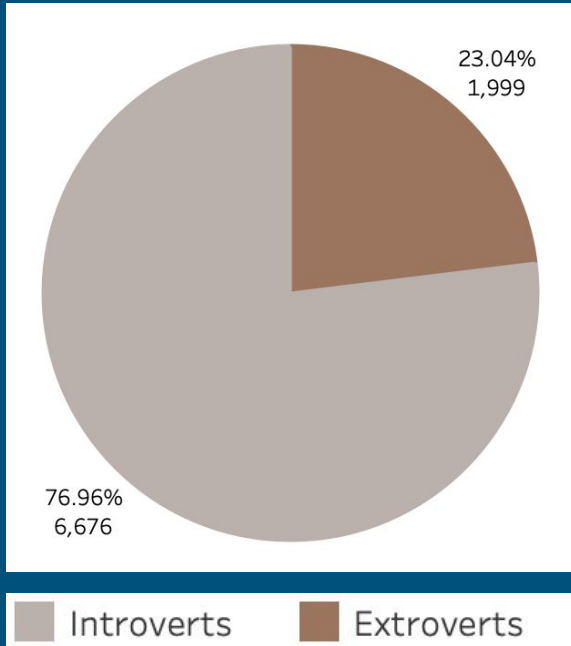
I

Introversion



Inside the Data:

Representation of E vs I



How do you prefer to take in information?

N

Intuition

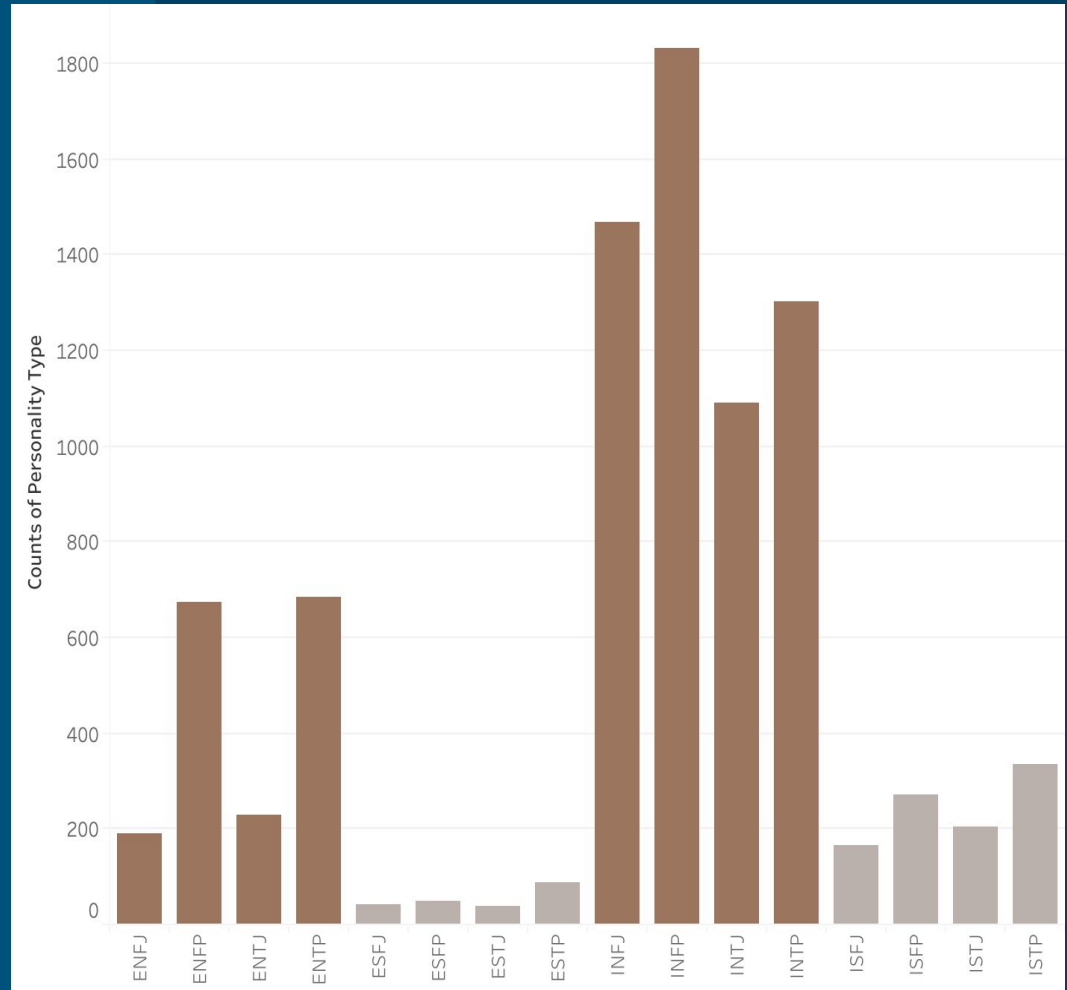
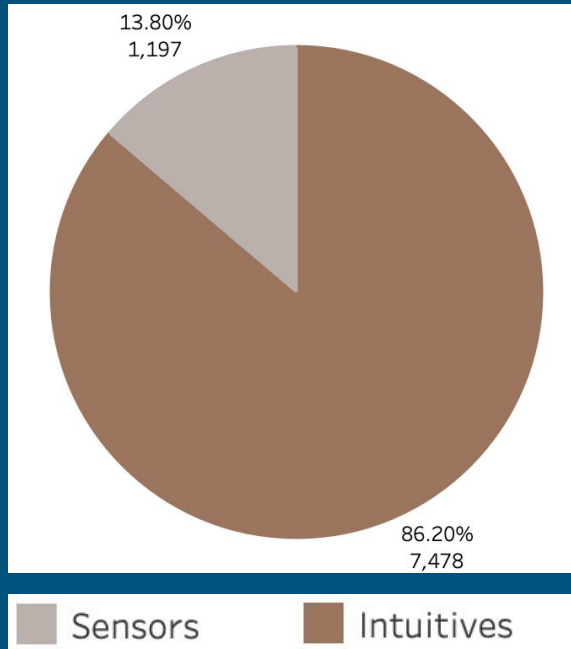
S

Sensing



Inside the Data:

Representation of **N** vs **S**



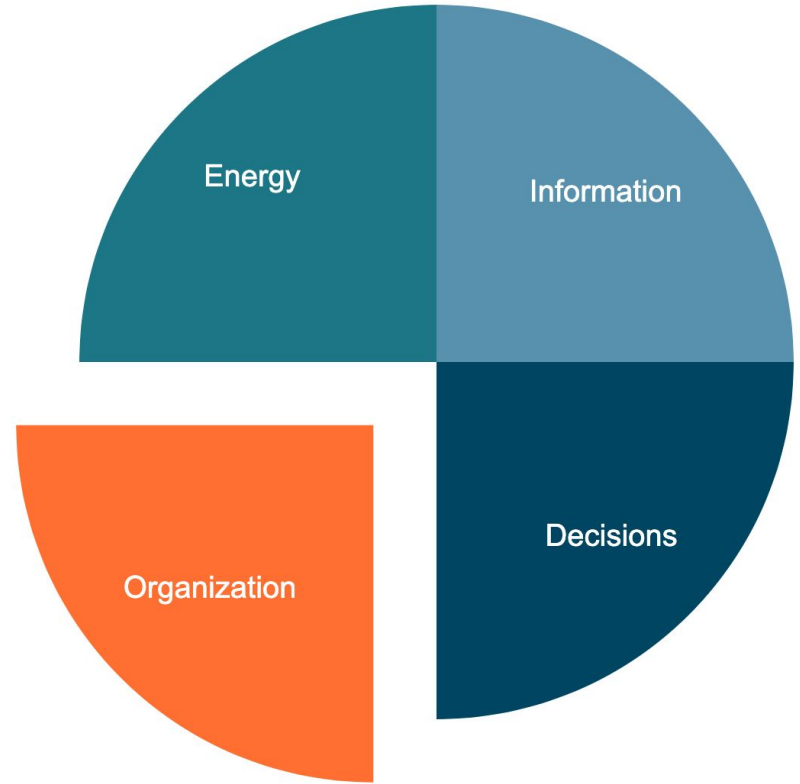
How do you prefer to make decisions?

F

Feeling

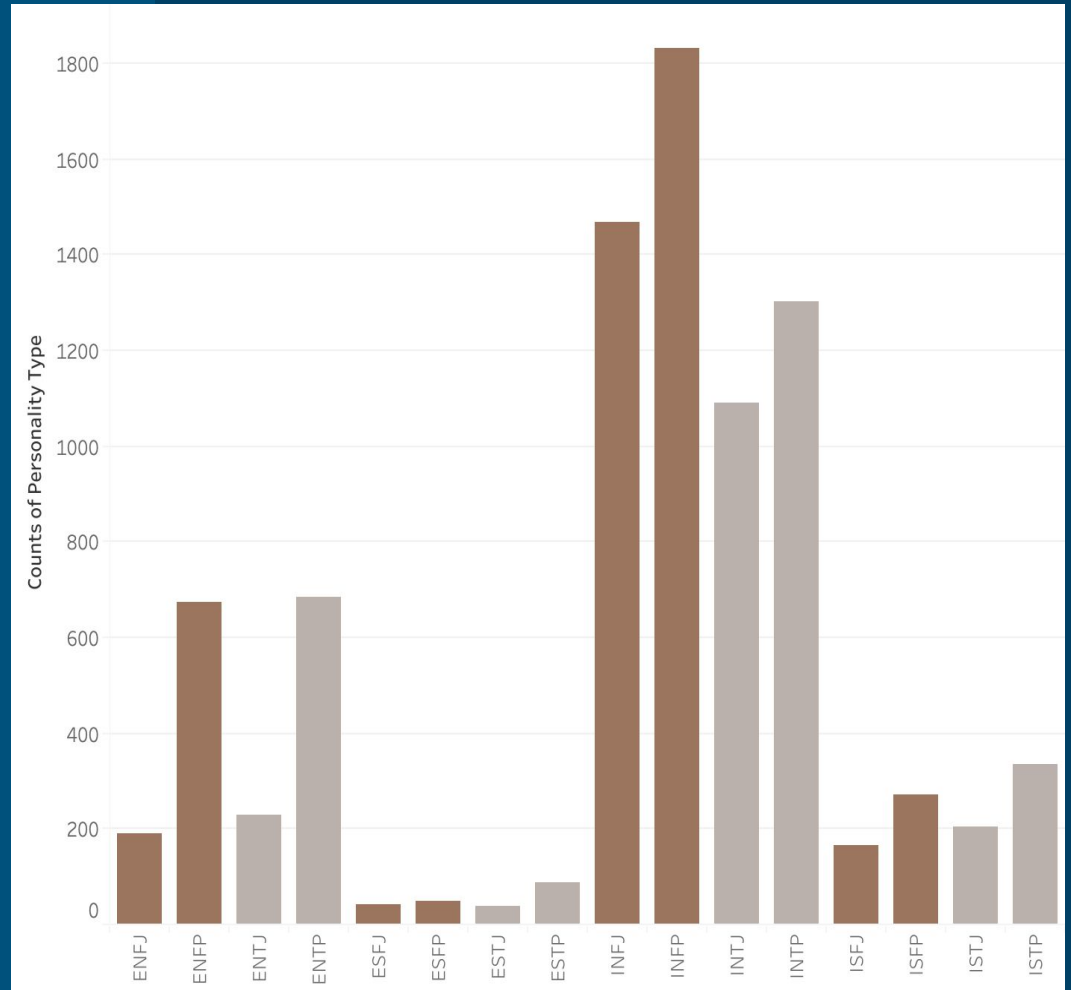
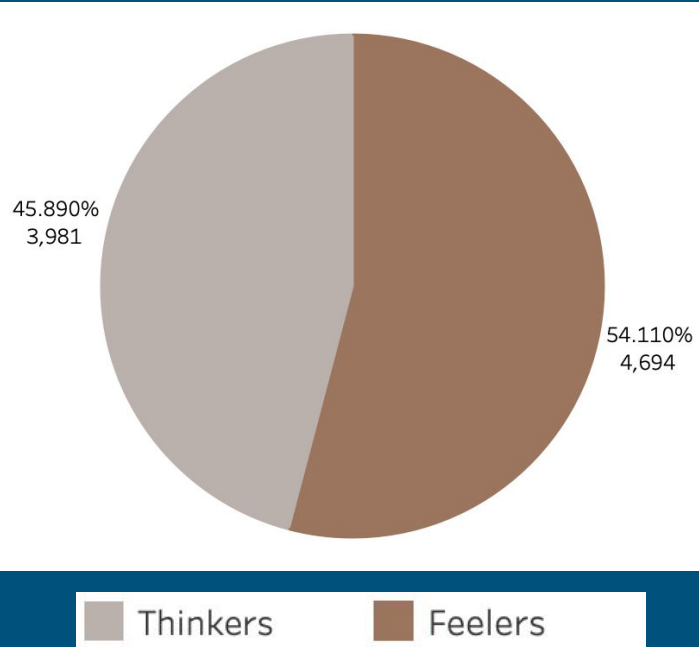
T

Thinking



Inside the Data:

Representation of F vs T



How do you prefer to live your outer life?

J

Judging

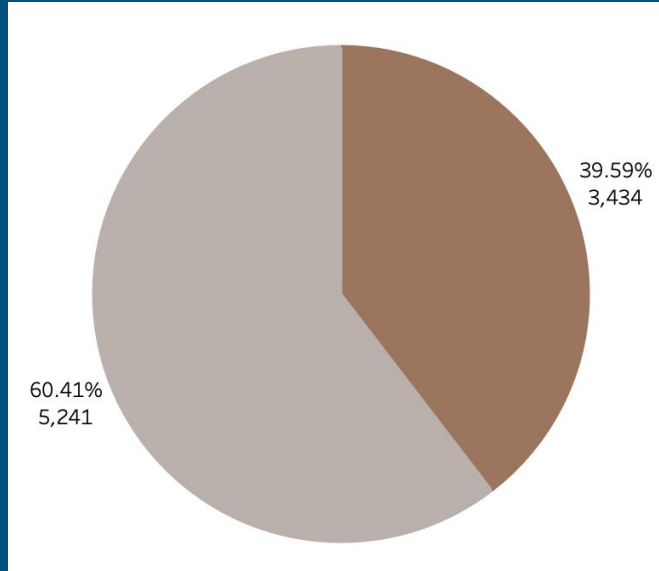
P

Perceiving

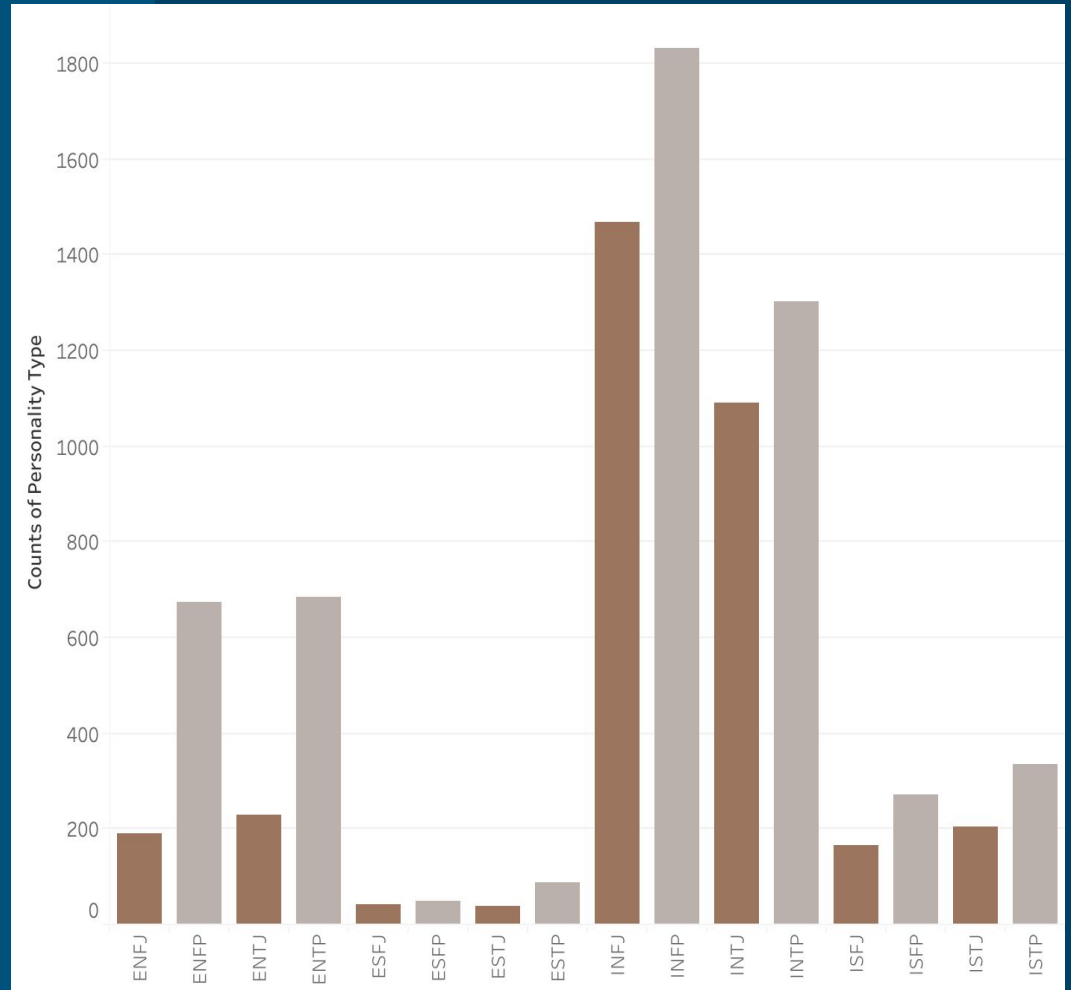


Inside the Data:

Representation of J vs P



Perceivers Judgers



Exploring the Posts

Post Text

Initial Observations

Many direct references to personality types.

Many references to websites.

Extensive and varied special and numeric character usage.

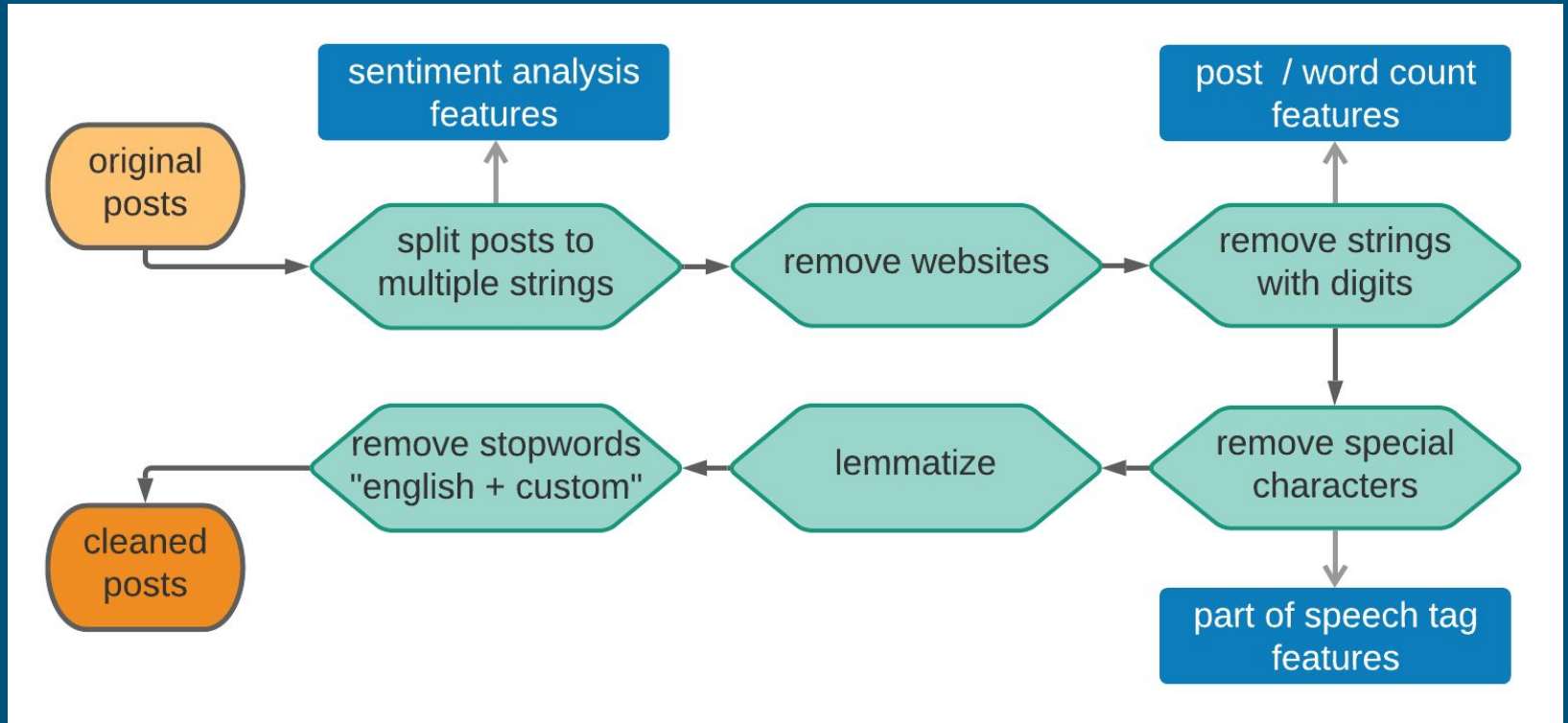
Exploring Features

Feature Engineering

Aside from the corpus of words,
what else can the text tell us?

- Sentiment Analysis
 - Average Words per Post
 - Number of Posts
- (excluding website references)*
- Part of Speech Composition Percentages
-

Cleaning the Text



Text Processing and Feature Engineering Diagram

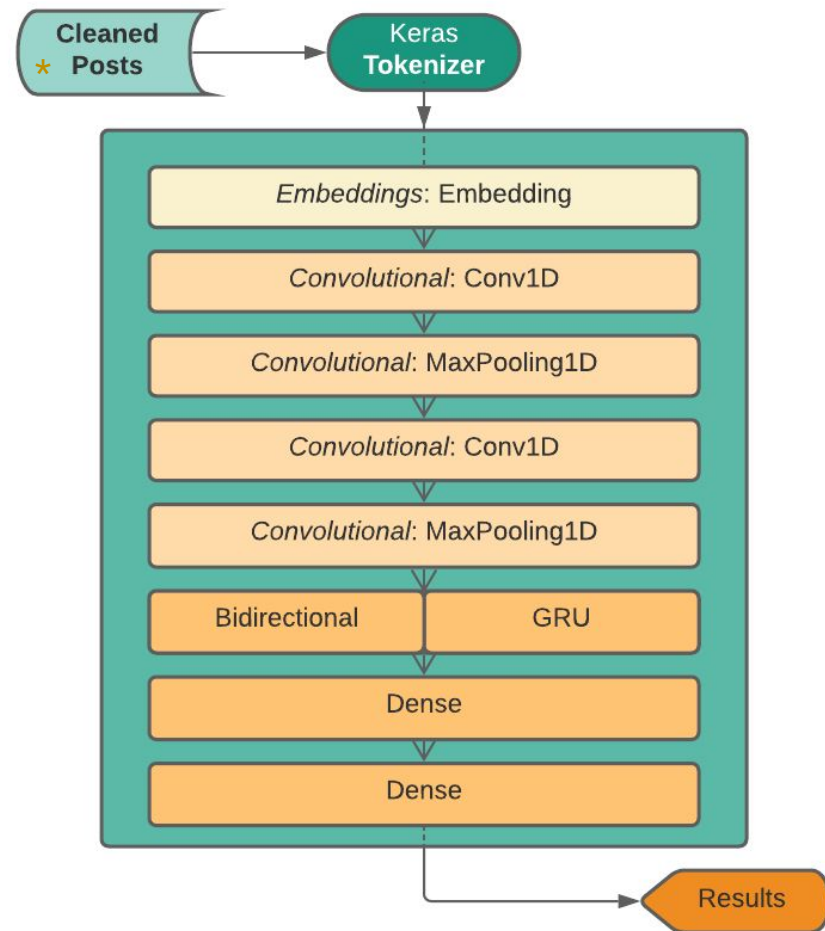
Building the Model

Recurrent Neural Network : **Architecture**

Keras Sequential Model

Overfitting Safeguards:

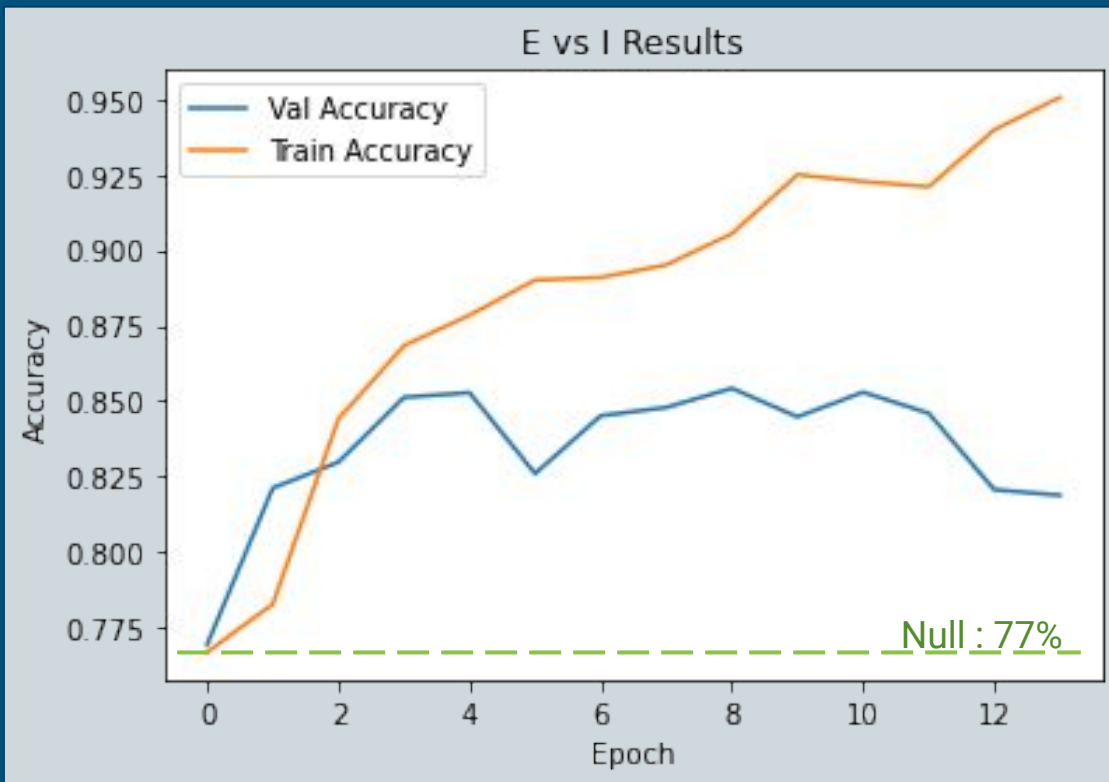
- Tokenizer Value Restriction
- Early Stopping
- Dropouts
- Recurrent Dropouts



Recurrent Neural Network : **Results**

Extroversion vs Introversion

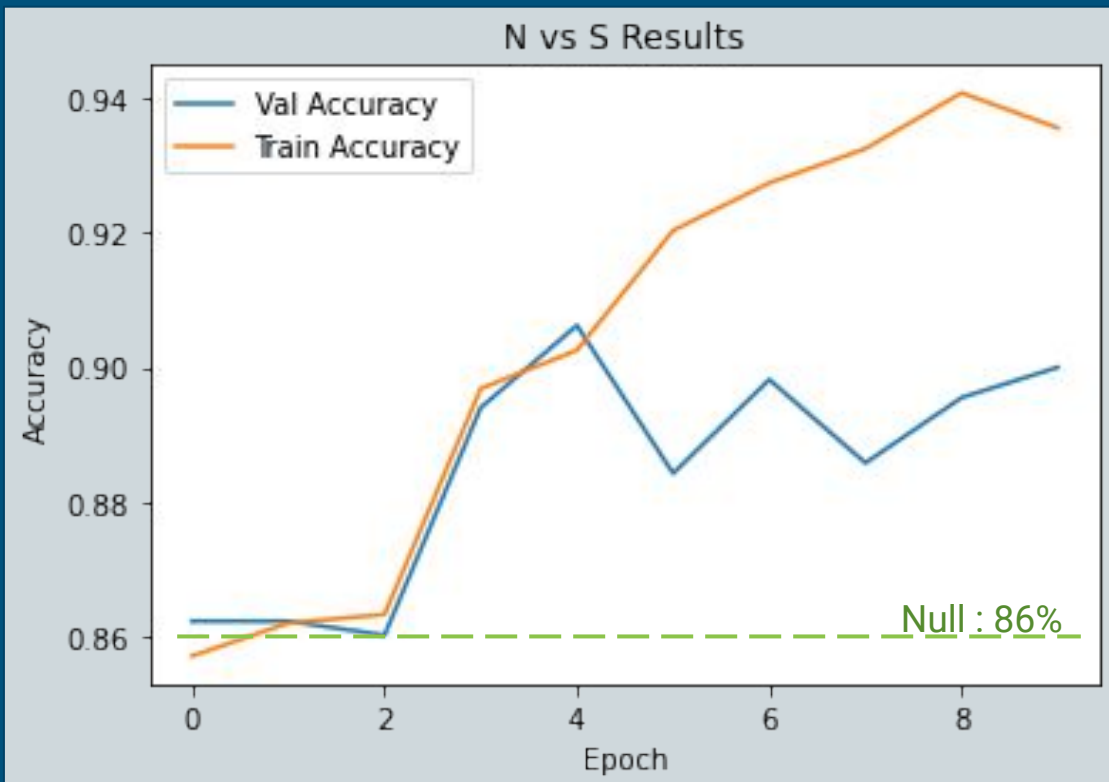
77% of pop.
Introverts



Recurrent Neural Network : **Results**

Intuitives vs Sensors

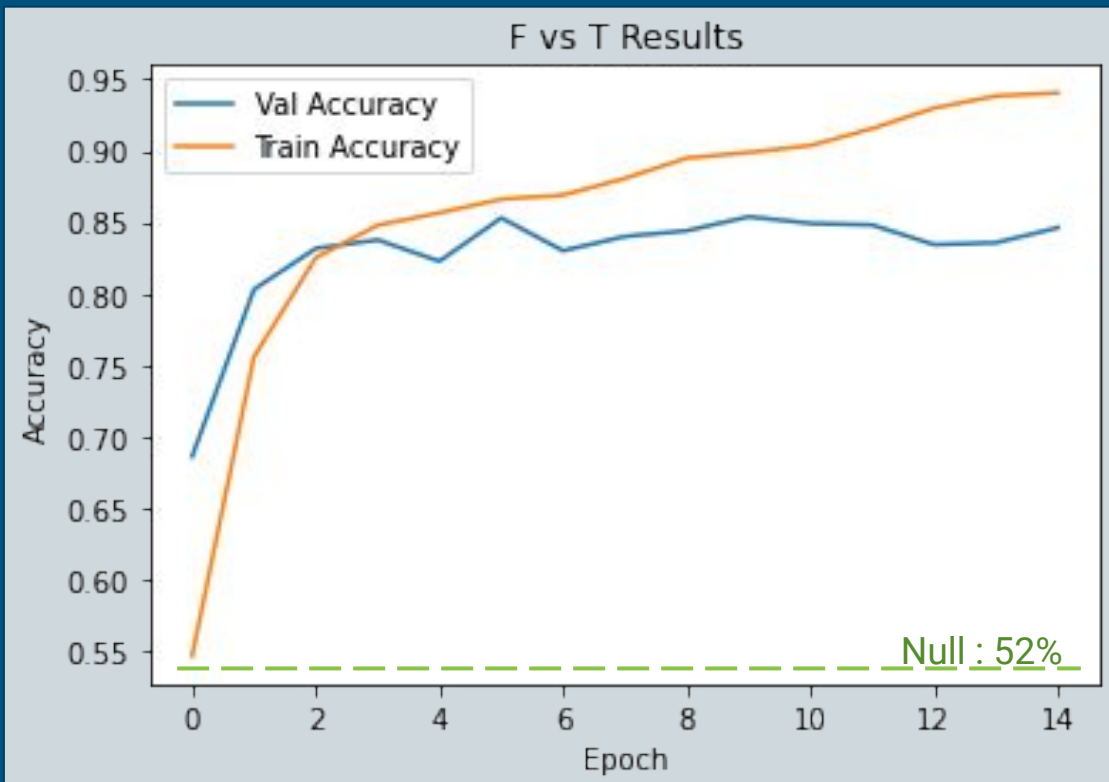
86% of pop.
Intuitives



Recurrent Neural Network : **Results**

Feelers vs Thinkers

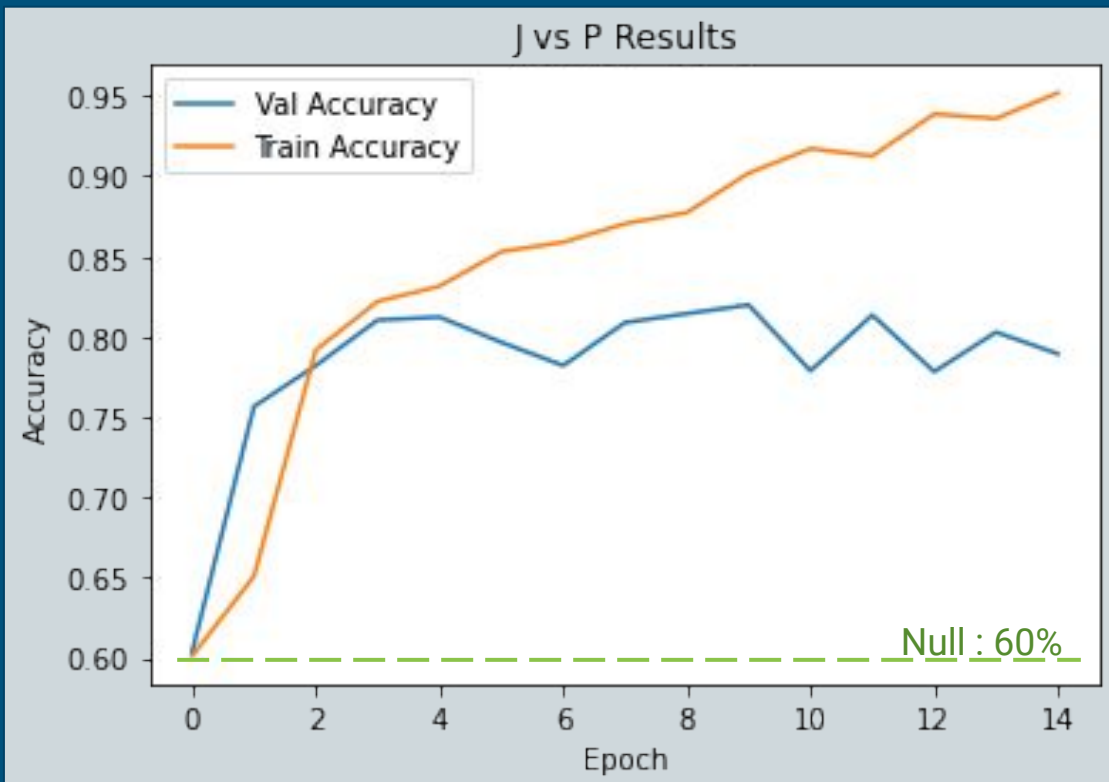
52% of pop.
Feelers



Recurrent Neural Network : **Results**

Judgers vs Perceivers

60% of pop.
Perceivers



Revisiting the Engineered Features

Engineered Features: Explored

Average Word Count

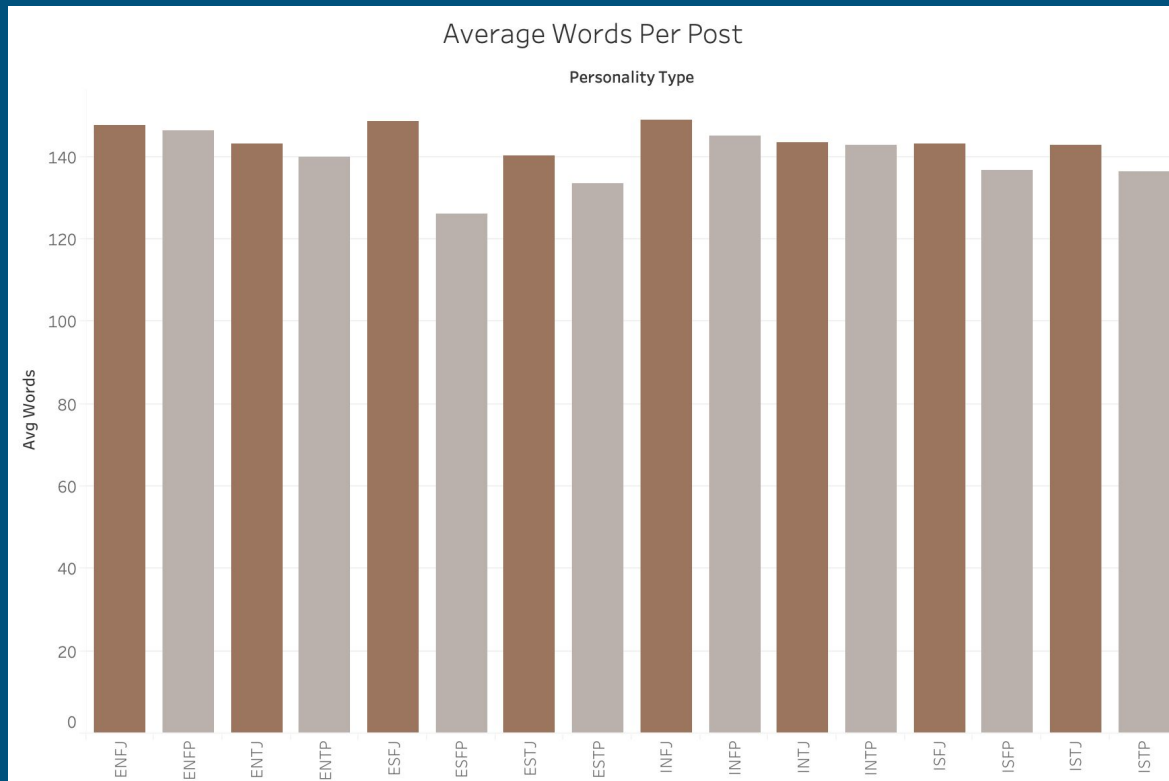
Judgers

saw consistently higher
average word counts over

Perceivers



analysis performed on cleaned post text

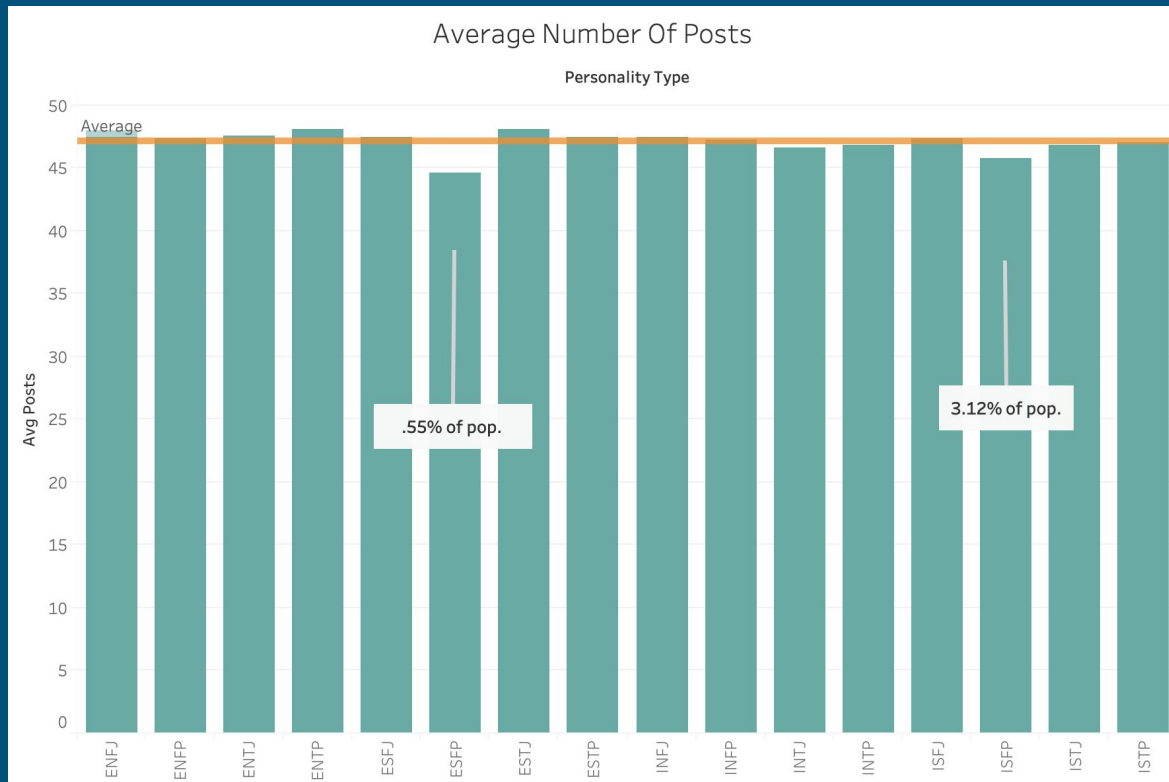


Engineered Features: Explored

Average Number of Posts

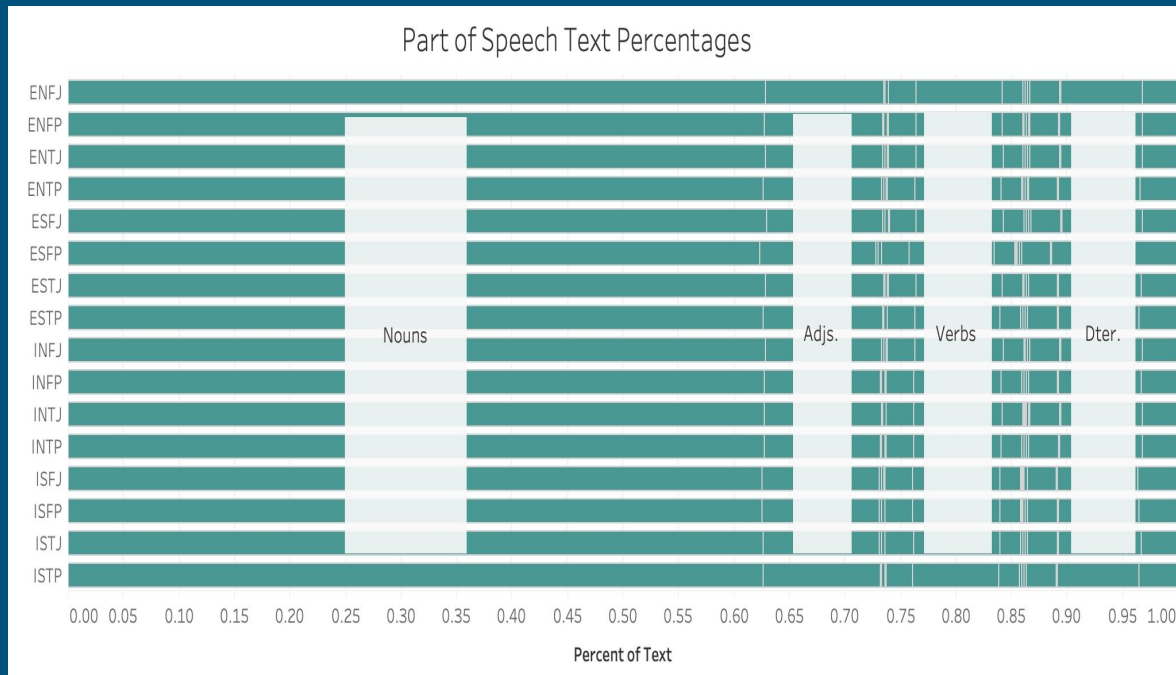
*Findings: Inconclusive
No Relationship Observed.*

★ analysis performed on cleaned post text



Engineered Features: Explored

Parts-of-Speech



*Findings: Inconclusive
No Relationship Observed.*

★ analysis performed on cleaned post text

Future Work

Repeat Neural Network Model

Extract Custom Stopwords

Custom Stopword Collection
Established:

- MBTI terms (eg. INTJ)
 - Enneagram terms (eg. si / sx)
 - Standard english terms
-

Revisit Supervised Learning Classifier Models

Pipeline Architecture Established:

- LogisticRegression
- LinearSupportVectorClassifier
- Gradient Boosting Classifier



